TEACHING ISSUES

TESOL Quarterly publishes brief commentaries on aspects of english language teaching.

Edited by DANA FERRIS
    University of California, Davis

Lexical Thresholds for Reading Comprehension: What They Are and How They Can Be Used for Teaching Purposes

BATIA LAUFER
University of Haifa
Haifa, Israel

■ We do not need to understand every single word when we read a text. Some words are not crucial for our comprehension and can therefore be ignored. Some unfamiliar words can be inferred from the text context. If they cannot be inferred, they can be looked up in a dictionary. However, ignoring unknown words, guessing them and looking them up in a dictionary cannot compensate for lack of good vocabulary knowledge. It is possible to ignore some unknown words in a novel, or a story, but not in texts dense in academic or professional information, such as medical or legal texts. Even if we can use contextual clues successfully, not all contexts provide clues for unknown words. Moreover, readers often ignore clues when they mistakenly consider unfamiliar words as familiar. Most importantly, clues often appear in words which themselves are unknown to learners and are therefore unusable. (For a discussion of inferring words from context, see Laufer, 1997, 2005.) A good dictionary is certainly helpful for finding the meaning of unknown words, assuming the reader knows how to use the dictionary efficiently. But looking up a large number of words may consume too much time and interfere with reading fluency.

Hence, neither guessing nor dictionary use strategies can compensate for insufficient knowledge of the text's vocabulary.

We can ask two related questions about the minimal lexis necessary for comprehension: (1) What percentage of a text's vocabulary should readers know to comprehend the text without resorting to compensatory strategies? (2) How large should the reader's vocabulary be in order to understand the necessary percentage of the text's words? The minimal vocabulary in terms of the percentage of familiar words in a text and the reader's vocabulary size is the lexical threshold required for understanding an authentic text.

Why is it important to quantify the lexical threshold? Most researchers agree that only when readers possess a critical mass of second language (L2) knowledge (lexical and grammatical) can general reading skills, such as distinguishing between main and peripheral information, between explicit and implicit material, operate most efficiently (Bernhardt & Kamil, 1995; Carrell, 1991; Clarke, 1980; Lee & Lemonnier-Schallert, 1997). Furthermore, vocabulary knowledge is a good predictor of reading proficiency, if not the best (Bernhardt & Kamil, 1995; Laufer, 1992; Nation 2001, 2006; Qian, 2002; Ulijn & Strother, 1990).

HOW MUCH OF A TEXT'S VOCABULARY SHOULD READERS KNOW TO REACH ADEQUATE COMPREHENSION?

Laufer (1989) found that the knowledge of 95% of the text's vocabulary was usually required to score 55% on a comprehension test and suggested that 95% of lexical coverage was the lexical threshold. (At the time of the study, 55% was the passing grade at her university.) In Hu & Nation (2000), *adequate* comprehension amounted to an average of 71% on two reading tests. The authors found that nobody could read adequately at 80% of coverage; some learners could at 90% and 95% coverage, but they were in the minority. The conclusion of the study was that 98% is the necessary lexical coverage for adequate comprehension. In a recent study, Schmitt, Grabe, and Jiang (2011) also suggested 98% as the optimal coverage for academic texts if a comprehension score of 70% is expected of learners. The two different coverage suggestions above, 95% and 98%, are the result of the two different reading scores considered to represent adequate comprehension.[1]

---

[1] I focus here on nonfiction written language only. For estimates of thresholds for spoken language, children's literature, and novels, see Nation (2006).

HOW LARGE SHOULD THE READERS' VOCABULARY BE IN ORDER TO UNDERSTAND THE NECESSARY PERCENTAGE OF THE TEXT'S WORDS?

In a large-scale study that included 745 participants, Laufer and Ravenhorst-Kalovski (2010) combined data on the lexical coverage of several academic texts, learners' vocabulary level, and reading comprehension scores of academic English. They found that learners whose vocabulary size was 6,000–8,000 word families reached lexical coverage of 98%; that is, they understood 98% of the text's words. Learners with vocabulary sizes of 4,000–5,000 word families reached 95% coverage. Using a corpus-based approach, Nation (2006) found that, in written texts, the 3,000 most frequent word families covered 89%–95% of the lexical items and 5,000 covered 92%–98% of the vocabulary. Proper nouns covered an additional 2%–4% of written texts. The converging results of the studies suggest that there are two thresholds for comprehending authentic written texts: an optimal one, which is the knowledge of 8,000 word families yielding the coverage of 98% (including proper nouns) and a minimal one, which is around 5,000, resulting in the coverage of 95% (including proper nouns).[2]

The findings presented here have important implications for setting vocabulary goals, designing lexical syllabi, and selecting materials and lexical tasks, particularly where reading is an important course component. Vocabulary size of the learners is crucial for comprehension, and teachers can easily measure it at the beginning and/or at the end of the course they teach, using one of the tests mentioned in note 2. The most recent test of vocabulary comprehension is a BNC (British National Corpus)–based vocabulary size test (Beglar, 2010; Nation & Beglar, 2007). According to the research results discussed earlier, we would, ideally, want learners to know 8,000 word families at the end of a reading course if we expect them to become independent readers of academic material. If learners reach the minimal threshold of 5,000 word families, they can be expected to manage the reading material with some assistance of a dictionary. If we know learners' vocabulary size at the beginning of their studies, we know what gap in lexical knowledge we have to bridge to attain one of the lexical thresholds. Thus, for example, learners whose vocabulary size is 3,000 word families will have to learn about 2,000 more from the fourth and fifth

---

[2] Some tests that can be used to measure how many word families (words with their inflections and common derivations) learners can understand are Vocabulary Levels Test (Nation, 1983), and its revised version (Schmitt, Schmitt, & Clapham, 2001), the Eurocentres vocabulary tests (Meara & Jones, 1989), and the recent Vocabulary Size Test (Nation & Beglar, 2007).

frequency levels if the goal of the course(s) is to help them read with support.

The words themselves can be found in various word frequency lists. Some of the lists are free. Three examples of such lists are: (1) the 5,000-word list based on COCA (Corpus of Contemporary American English) that can be downloaded from Mark Davies and Dee Gardner's site (http://www.wordfrequency.info/5k_lemmas_download. asp); (2) 20 frequency lists (from K1, the first 1,000 most frequent words to K20); and (3) lists of the most frequent 2000 words together with the AWL (Academic Vocabulary List), both available at Lextutor, Tom Cobb's website. For the BNC lists, see http://conc.lextutor.ca/list_learn/bnc/. For 2K + AWL, see http://www.lextutor.ca/freq/lists_download/. The AWL is particularly useful in courses of English for Academic Purposes because the 2,000 most frequent words and the AWL together, a combined list of 2,570 words, can bring the coverage of an academic text up to approximately 90%.

The lists of unfamiliar vocabulary constitute a good source for word-focused instruction, or exercises that are specifically designed for practicing selected words. Word-focused instruction compensates for the paucity of word appearances in the input. Words that are less common than the 2,000 most frequent vocabulary items cannot be expected to reappear in the texts often enough for learners to notice or remember them from input only. Hence learners need specific activities, exercises, and tests that draw particular attention to the new vocabulary.

Teachers can easily find out how difficult a text is for their students, lexically, by analyzing the text in terms of lexical frequency levels. Such analyses are computerized and freely available at Lextutor, at http://www.lextutor.ca/ (Go to Vocabprofile or Range). The BNC-based vocabulary profile analysis shows what percentage of words in the text belong to the first 1,000 most frequent words, what percentage belong to the second 1,000, third, fourth, and so on, until the 20th thousand. An older version of the profile analyzes a text into the first 1,000, second 1,000, academic vocabulary, and all the other words that do not belong to these three categories. For example, a vocabulary profile can show that 80% of the text's words are from the first 1,000, 10% from the second 1,000, and 5% from the third. This means that 3,000 words provide 95% coverage of the text's vocabulary. If learners' vocabulary size is found to be 3,000 words, this will mean that they can read the text with support. The remaining less frequent unfamiliar words that constitute 5% of the entire text are candidates for learning. We can "visualize" the notion of lexical coverage by imagining a written page, which is approximately 300 words, with 15 (5%) unknown words. As mentioned earlier, learners can comprehend such a text with a dictionary and some inferring from context. A 98%

lexical coverage would mean that there are 6 unknown words on the page (2% out of 300); 90% coverage would mean 30 unknown words. The former coverage would result in fluent reading and good comprehension. Researchers recommend it for "reading for pleasure." As for the latter example, though it is possible to comprehend such a text with appropriate dictionary use, the flow of reading will be interrupted too often. Hence it is questionable whether a tedious and probably frustrating reading activity is effective for learning.

In this article, I discussed the notion of lexical threshold in terms of the percentage of words learners need to understand in a text and learners' vocabulary size. I argued that a text's lexical profile and learners' vocabulary size, for which quantitative measures are available, provide teachers with the necessary information about the lexical ease and difficulty of different texts for different learners. They also show which words are likely to be beyond learners' knowledge and are, therefore, potentially good candidates for word-focused instruction.

THE AUTHOR

Batia Laufer is professor of applied linguistics in the Department of English Language and Literature at the University of Haifa, Israel. Her main research interests are vocabulary acquisition in additional languages, lexicography, cross-linguistic influence, and language attrition.

REFERENCES

Beglar, D. (2010). A Rasch-based validation of the vocabulary size test. *Language Testing*, *2*, 101–118. doi:10.1177/0265532209340194

Bernhardt, E. B., & Kamil, M. L. (1995). Interpreting relationships between L1 and L2 reading: Consolidating the linguistic threshold and the linguistic interdependence hypotheses. *Applied Linguistics*, *16*, 15–34. doi:10.1093/applin/16.1.15

Carrell, P. L. (1991). Second language reading: Reading ability or language proficiency? *Applied Linguistics*, *12*, 159–179. doi:10.1093/applin/12.2.159

Clarke, M. A. (1980). The "short-circuit" hypothesis of ESL reading—Or when language competence interferes with reading performance. *Modern Language Journal*, *64*, 203–209. doi:10.1111/j.1540-4781.1980.tb05186.x

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*, 403–430. Retrieved from http://nflrc.hawaii.edu/rfl/

Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Clevedon, England: Multilingual Matters.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In H. Bejoint & P. Arnaud (Eds.), *Vocabulary and applied linguistics* (pp. 126–132). Basingstoke, England: Macmillan.

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 20–34). Cambridge, England: Cambridge University Press.

Laufer, B. (2005). Instructed second language vocabulary learning: The fault in the 'default hypothesis.' In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 286–303). Berlin: Mouton de Gruyter.

Laufer, B., & Ravenhorst-Kalovski, G. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22* (1), 15–30. Retrieved from http://nflrc.hawaii.edu/rfl/

Lee, J. L., & Lemonnier-Schallert, D. (1997). The relative contribution of L2 language proficiency and L1 reading ability to L2 reading performance: A test of the threshold hypothesis in an EFL context. *TESOL Quarterly, 31,* 713–739. doi:10.2307/3587757

Meara, P. M., & Jones, G. (1989). *Eurocentres Vocabulary Test 10 KA.* Zurich, Switzerland: Eurocentres.

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines, 5,* 12–25.

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge, England: Cambridge University Press.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63,* 59–82. doi:10.1353/cml.2006.0049

Nation, I. S. P., & Beglar, D. (2007) A vocabulary size test. *The Language Teacher, 31*(7), 9–13. Retrieved from http://jalt-publications.org/tlt

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52,* 513–536. doi:10.1111/1467-9922.00193

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18,* 55–88. doi:10.1177/026553220101800103

Schmitt, N., Grabe, W., & Jiang, X. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95,* 26–43. doi:10.1111/j.1540-4781.2011.01146.x

Ulijn, J. M., & Strother, J. B. (1990). The effect of syntactic simplification on reading EST texts as L1 and L2. *Journal of Research in Reading, 13,* 38–54. doi:10.1111/j.1467-9817.1990.tb00321.x